Model-Based Quantification of Image Quality

Rajeeb Hazra, Keith W. Miller, Stephen K. Park

Introduction [1]

In 1982, Park and Schowengerdt [1] published an end-to-end analysis of a digital imaging system quantifying three principal degradation components (i) *image blur* - blurring caused by the acquisition system (ii) *aliasing* - caused by insufficient sampling and (iii) *reconstruction blur* - blurring caused by the imperfect interpolative reconstruction. This analysis, which measures degradation as the square of the radiometric error, includes the sample-scene phase as an explicit random parameter and characterizes the image degradation caused by imperfect acquisition and reconstruction together with the effects of undersampling and random sample-scene phases. In a recent paper Mitchell and Netravelli [3] displayed the visual effects of the above mentioned degradations and presented subjective analysis about their relative importance in determining image quality.

The primary aim of the research in this paper is to use the analysis of Park and Schowengerdt [1],[8] to correlate their *mathematical* criteria for measuring image degradations with subjective *visual* criteria. Insight gained from this research can be exploited in the end-to-end design of optical systems, so that system parameters (transfer functions of the acquisition and display systems) can be designed relative to each other, to obtain the "best possible" results using quantitative measurements.

Formulation

In this section we present an end-to-end model of a digital imaging system. This model was used by Park and Schowengerdt [1] to derive expressions for the degradation caused by the various components of the system.

The model upon which the results of this paper are based is described in Fig 1. The parameters u and v are explicit sample scene phase parameters which have the range of $\pm\frac{1}{2}$ for pixels placed at unit distance from each other. The action of the imaging subsystem is described by the convolution (denoted by $*$) of the system point spread function (PSF) $h(x,y)$ with the scene

$$g(x-u, y-v) \;=\; h(x,y) * f(x-u, y-v) \tag{1}$$

The image is then sampled onto a cartesian grid. This sampling operation is represented symbolically as the multiplication of the image with the *comb* or *Shah* function

$$g_s(x,y;u,v) = \sum_m \sum_n g(x-u, y-v)\delta(x-m, y-n) \tag{2}$$

The notation $g_s(x,y;u,v)$ expresses the fact that the sampling subsystem is not shift-invariant.

[1] This paper refers to research described in references 1-8.

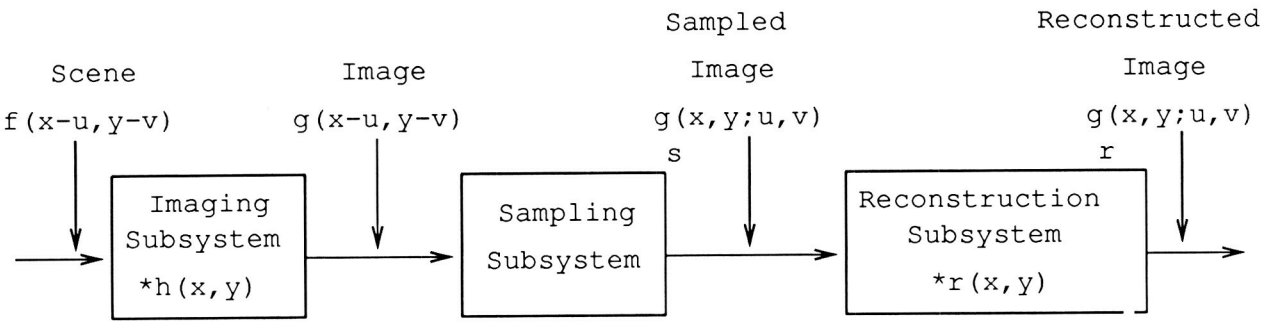

Figure 1: An Imaging, Sampling and Reconstruction System

We take the point of view that the reconstruction filter $r(x, y)$ is designed so that the reconstructed image is an accurate reproduction of the output of the imaging system. The reconstructed image is compared to the image g and not the scene f; thus, the reconstruction filter typically does not attempt to perform any restoration.

Reconstruction is also symbolically modeled as a convolution operation

$$g_r(x, y; u, v) = r(x, y) * g_s(x, y; u, v) \tag{3}$$

Park and Schowengerdt measure the *accuracy* of reconstruction as the mean square radiometric error and define the term

$$\epsilon^2_{SR}(u, v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [g(x - u, y - v) - g_r(x, y; u, v)]^2 dxdy \tag{4}$$

and analogously

$$\epsilon^2_I = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [f(x - u, y - v) - g(x - u, y - u)]^2 dxdy \tag{5}$$

where ϵ^2_{SR} and ϵ^2_I measure the *sampling-reconstruction* degradation and *image blur* respectively. As suggested by the notation, image blur is independent of the sample scene phase due to the shift invariance of the convolution operation. The sampling-reconstruction degradation is not.

Fourier analysis yields equivalent expressions for ϵ^2_{SR} and ϵ^2_I in the frequency domain. Park et al. showed that

$$\epsilon^2_I = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |\, 1 - \hat{h}(\nu_x, \nu_y)\,|^2 |\, \hat{f}(\nu_x, \nu_y)\,|^2 \, d\nu_x d\nu_y \tag{6}$$

where (ν_x, ν_y) are spatial frequencies (units of cycles per sampling interval), $\hat{h}(\nu_x, \nu_y)$ is the imaging subsystem OTF (optical transfer function) and $|\, \hat{f}(\nu_x, \nu_y)\,|$ is the magnitude of the transform of the scene.

The corresponding expression for the sampling and reconstruction degradation is given in terms of an ensemble of scenes formed by varying the sample scene phase parameters uniformly over their entire range. Thus, we obtain the *expected value* of this degradation in the form

$$E[\epsilon^2_{SR}] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^2(\nu_x, \nu_y) \,|\, \hat{h}(\nu_x, \nu_y)\,|^2 |\, \hat{f}(\nu_x, \nu_y)\,|^2 \, d\nu_x d\nu_y \tag{7}$$

where the term $e^2(\nu_x, \nu_y)$ accounts for the effects of imperfect reconstruction and undersampling and is given by

$$e^2(\nu_x, \nu_y) = \mid 1 - \hat{r}(\nu_x, \nu_y) \mid^2 + \sum_{(m,n) \neq (0,0)} \sum \mid \hat{r}(\nu_x - m, \nu_y - n) \mid^2 \tag{8}$$

where $\hat{r}(\nu_x, \nu_y)$ is the reconstruction filter, i.e. the Fourier transform of $r(x, y)$. $E[\epsilon_{SR}^2]$ can be written as the sum of two terms,

$$E[\epsilon_{SR}^2] = \epsilon_R^2 + \epsilon_S^2 \tag{9}$$

where

$$\epsilon_R^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mid 1 - \hat{r}(\nu_x, \nu_y) \mid^2 \mid \hat{h}(\nu_x, \nu_y) \hat{f}(\nu_x, \nu_y) \mid^2 \, d\nu_x d\nu_y \tag{10}$$

and

$$\epsilon_S^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\sum_{m,n \neq 0,0} \sum \mid \hat{r}(\nu_x - m, \nu_y - n) \mid^2 \right] \mid \hat{h}(\nu_x, \nu_y) \hat{f}(\nu_x, \nu_y) \mid^2 \, d\nu_x d\nu_y \tag{11}$$

The term ϵ_R^2 accounts for imperfect reconstruction while ϵ_S^2 accounts for aliasing due to undersampling.

Analysis and Visual Perception of Image and SR Blur

Image blur is caused by the non-ideal frequency response of the imaging subsystem. Eq. (6) is a mathematical statement of this fact. Almost invariably, the frequency response of an imaging system approaches zero at high frequencies and thus this subsystem acts as a low-pass filter. Image blur alone, uncoupled from sampling and reconstruction blur, is perceived as a loss of high frequency detail in the scene.

The average sampling and reconstruction blur, as suggested by Eq. (7) is caused by inadequacies in both the sampling and the reconstruction subsystem. The sampling contribution to this degradation is expressed by Eq. (11) which states that aliasing is caused by the presence of significant image energy at frequencies where the energy in the reconstruction filter sidebands

$$\sum_{m,n \neq 0,0} \sum \mid \hat{r}(\nu_x - m, \nu_y - n) \mid^2 \tag{12}$$

is not zero. This is illustrated in Fig 2 where the replicas of the image spectrum (formed by sampling) overlap and the reconstruction filter cannot isolate a pure version of the *base-band* spectrum. This type of degradation is sometimes called *prealiasing* [3] and will always be present if the image is not sufficiently sampled, even with perfect reconstruction.

Even when the replicated spectra do not overlap (i.e the image has been sufficiently sampled), image quality may suffer due to poor reconstruction, as illustrated in Fig 3. In this case, the response of the reconstruction filter is too broad and thus the reconstructed signal includes some (high) frequencies not present in the original image. This type of aliasing is sometimes called *postaliasing* [3]. When the image spectrum has significant power at frequencies very near the Nyquist (cutoff) frequency (i.e the image spectrum and its nearest replica come very close to each other), the design of the reconstruction filter becomes difficult as the roll-off has to be very sharp (resulting in a filter with a very large kernel in the spatial domain). This problem has been noted by several researchers [2], [3].

Pre- and postaliasing are often perceived as artifacts in the reconstructed scene [3]. However, it should be noted that in general, an absence of artifacts does not imply that there is no pre or postaliasing. Aliasing can manifest itself as blurring as well (due to attenuation of the high frequencies in the scene or image spectrum) and is almost impossible to differentiate from image blur.

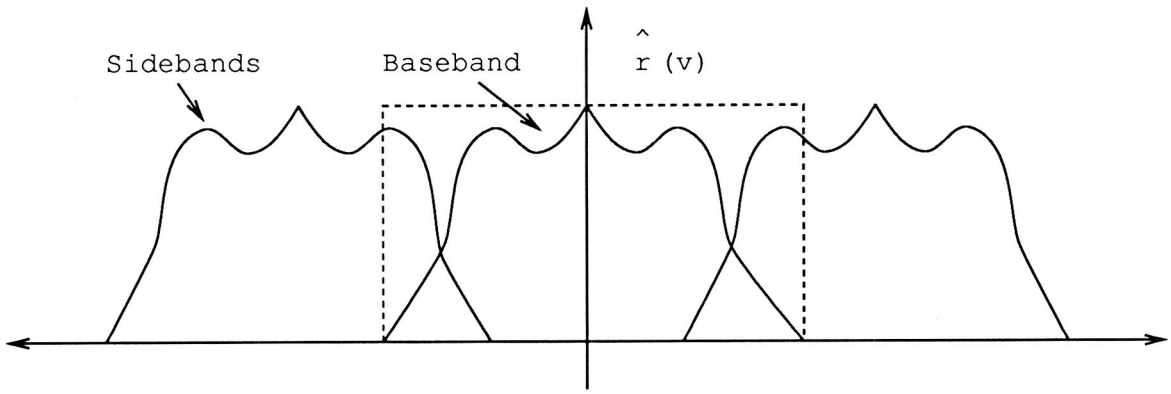

Figure 2: Prealiasing resulting from insufficient sampling

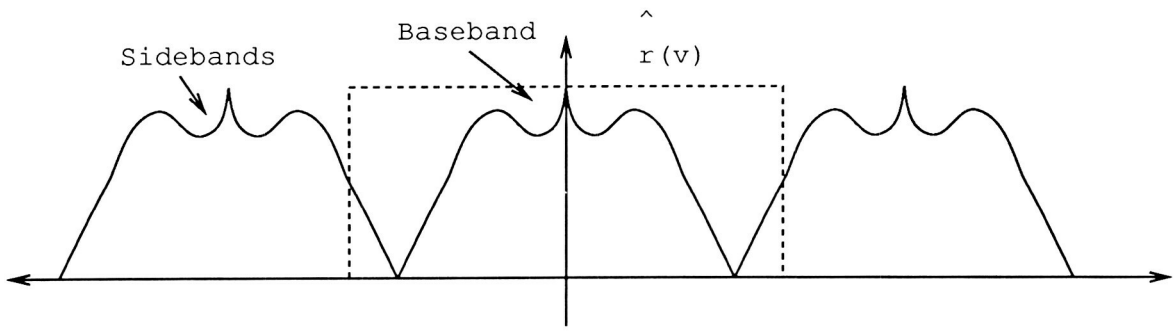

Figure 3: Postaliasing resulting from poor reconstruction

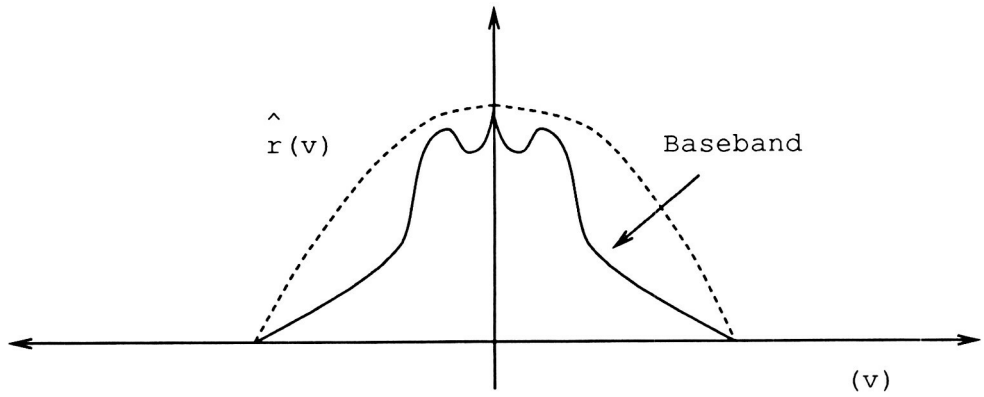

Figure 4: Baseband attenuation resulting from imperfect reconstruction

In addition to removing the sidebands of the signal spectrum, the reconstruction filter also needs to pass the original image spectrum base-band with minimal distortion (Fig 4). Eq. (10) states this idea formally. It measures the contribution to SR blur caused by the presence of significant image energy, $|\hat{h}(\nu_x,\nu_y)\hat{f}(\nu_x,\nu_y)|^2$ at frequencies where $\hat{r}(\nu_x,\nu_y) \neq 1$. This type of reconstruction error is known as *base-band attenuation.* This is analogous to image blur in the sense that the reconstruction filter acts as a low-pass filter resulting in a loss of high-frequency detail in the reconstructed image.

The problem of designing a good reconstruction filter is made difficult because an "ideal" filter is a sinc filter in the spatial domain. The sinc is quasi-ideal in the sense that a signal can be *perfectly* reconstructed from its samples by using sinc-interpolation only if the signal is bandlimited and sufficiently sampled. However, the sinc is impossible to realize in practice and finite approximations to it produce an effect commonly known as *ringing.* Ringing is perceived as rippling patterns radiating from high contrast edges [3] and is strongly suggested by the form of the impulse response of the sinc.

Another problem in designing a reconstruction filter is the problem of *sample-frequency ripple.* This problem can be best understood in terms of a uniformly gray image which is sampled and reconstructed to yield an image where the gray-level uniformity is destroyed. This is often perceived as spurious patterns on the background in an image. To eliminate this problem, it is necessary to design the reconstruction filter $\hat{r}(\nu_x,\nu_y)$ so that the equation

$$\sum_m \sum_n r(x-m, y-n) = 1 \tag{13}$$

is satisfied.

An important point to note in this discussion is that even though it is possible (at least in theory) to minimize image blur and sampling-reconstruction blur individually by suitable filter design, in an end-to-end system the subsystems cannot be designed in isolation from one another to minimize both image and sampling-reconstruction blur simultaneously. Eq. (6) suggests that image blur is minimized when $\hat{h}(\nu_x,\nu_y) = 1$ for all frequencies where there is non-zero scene energy. However, from Eq. (8) we see that the average sampling-reconstruction blur will be minimized when $\hat{h}(\nu_x,\nu_y) = 0$ at all those frequencies where the reconstruction filter does not have unit (perfect) response. These are conflicting requirements and a compromise has to be achieved based on the *relative visual importance* of the two types of degradation.

It has been observed [6], [7] that the response of human viewers to various spatial effects of filters is subjective. Filters that result in some aliasing and base-band attenuation have sometimes been observed to yield results which are visually pleasing to human viewers. There is evidence [6] that suggests that a moderate amount of ringing can "improve" the visual quality of an image by introducing an illusion of sharpness (high frequency), although in terms of the amount of degradation (which can be measured by Eq. 6 - 11), this may correspond to a higher mean squared error.

In our research, we attempt to correlate the mathematical criteria for optimal end-to-end processing with subjective visual testing for Gaussian transfer functions. There is evidence that people prefer some aliasing and ringing (which give an illusion of sharpness), but that people are sensitive to high-frequency suppression (blurring). The primary motivation of this study is to assess the effect of each of these individual degradation components on the quality of the reconstructed image. This research has application to the design of end-to-end imaging systems where the components can be tuned to obtain the best possible results. In the next section we describe our models for the various components of the system and simulation results.

Imaging and Reconstruction System Transfer Function Models

In order to simulate an end-to-end imaging system it is necessary to associate a model with the imaging (camera) subsystem and the reconstruction (display) subsystem;i.e,we need to assign a *functional* form to both $\hat{h}(\nu_x, \nu_y)$ and $\hat{r}(\nu_x, \nu_y)$. In the discussion that follows, we refer to $\hat{h}$ as the Camera Transfer Function (CTF) and $\hat{r}$ as the Display Transfer Function (DTF). In our analysis we have chosen to model the CTF and DTF as Gaussian functions of the form

$$e^{-\left[\frac{1}{r^2}(\nu_x^2+\nu_y^2)\right]} \tag{14}$$

where r is the parameter which controls the spread of the function. It can be shown that r is proportional to the standard deviation of the Gaussian function (σ_ν). They are related as

$$r = \sqrt{2}\sigma_\nu \tag{15}$$

The two-dimensional Gaussian function is separable and its Fourier transform is also a Gaussian.

In our model, the Gaussian is symmetric in the two dimensions resulting in the filter kernels being circularly symmetric. Thus, only a single parameter (r) is required to characterize each of the Gaussian functions representing the CTF and the DTF. Thus, due to the duality of the Gaussian and its Fourier transform, a broad frequency response can be achieved by a very small kernel in the spatial domain and vice versa.

The reason for modelling the CTF and the DTF as Gaussian functions is primarily due to the its popularity amongst designers of optical instruments [4], [5]. Several variations of the pure Gaussian (e.g. sharpened Gaussian and the sum of two Gaussians [4]) have been used as models for the transfer functions, especially for interpolative reconstruction systems. In our end-to-end model, we have three system parameters - the sampling rate and the standard deviations of the Gaussian camera and display transfer functions. These parameters can be varied to influence both image and sampling and reconstruction blur. End-to-end simulation using these models for the imaging and reconstruction subsystems thus allows us to study the interplay between the various degradations discussed in the previous section and correlate mathematical results (blur coefficients) with subjective (visual) judgements about image quality. The primary goal of this simulation study is to identify a relationship between the two parameters which will result in the best possible reconstructed image. Such a relationship can then be used as a design rule for end-to-end systems employing scanning and interpolative reconstruction.

Simulation Results

The numerical simulation of the imaging system described in Fig 1. has been performed on two images - one of cat's face and the other of the central portion of a dollar bill. The images are 512 x 512 pixels in dimension and are quantized to 16 bits/gray-level. For the purpose of display, these images have been rescaled into a gray-level range of 0 to 255 (8 bits/gray-level).

In the simulation of the imaging process, the Fourier spectra of these scenes have been multiplied with a Gaussian CTF to produce the corresponding image spectra. The reverse Fourier transform of the image spectra produces the corresponding image in the spatial domain. It is with this image that the final reconstructed scene is compared to judge the quality of reconstruction.

The sampling subsystem has been simulated by sub-sampling the 512 x 512 image down to 128 x 128. A uniform sampling scheme has been chosen primarily due to its simplicity as well as its popularity amongst designers of (digital) optical equipment.

Finally, the reconstruction process is implemented in a manner similar to the imaging process. The sampled images have been enlarged to 512 x 512 by zero-filling and their Fourier spectra have then been multiplied with a Gaussian DTF to produce the spectra of the reconstructed scenes. The inverse Fourier transform is then applied to these spectra to produce the reconstructed scenes in the spatial domain.

The parameter of interest for the transfer functions is r (Eq. 14.), which controls the standard deviation of the Gaussian functions used to model the CTF and the DTF. In order to study the degradation caused by these two subsytems, r_{CTF} and r_{DTF} have been varied over a range of values - the range selected is standardized with respect to the size of the sampled image (128 x 128).

The reconstructed scenes have been evaluated by about 20 people and the degradation values corresponding to their choice of the best possible reconstruction are shown on the corresponding plots. The observers were first shown the images after they were passed through the imaging subsystem and were then asked to find the most faithful reconstruction from the collection of processed images for different system parameters. The candidate images were displayed in a random order to eliminate any positional bias that may have been present. The contrasts of these images were also strictly matched to eliminate any contrast bias.

Fig 5 shows the cat image after being passed through the acquisition phase of an imaging system. Figs 6 - 8 show reconstructed images of this acquired image with different display subsystems. Figs 9 - 14 are plots of the different error components for the dollar and the cat images. These values have been calculated using Eq.(6)-(11) and the horizontal axis of the plots refer to a fraction of 128 which represents the value of r_{CTF}. Figs 15 - 17 show several processed versions of the dollar bill image. Fig 15 shows the original dollar bill which serves as the scene in our simulations. Fig 16 and Fig 17 show results from two opposite ends of the processing spectrum - Fig 16. shows the excessive blurring introduced by the narrow (frequency domain) transfer functions, while Fig 17 exhibits the characteristic sample-frequency ripple associated with a wide (frequency domain) display transfer function.

The preliminary results of the visual testing have yielded interesting "observations" about the visual impact of the different kinds of degradations that are inevitably introduced in a nonideal end-to-end imaging system; 18 out of the 20 observers chose the reconstructed image corresponding to $r_{CTF} = 76.8$ (i.e. $CTF\ FACTOR = 0.6$ in Fig 7) as the "best" reconstructed scene for the cat image. From the plot it is clear that these values of the parameters correspond to a situation where the total degradation is dominated by the sampling (or aliasing) blur. This reinforces the belief that the human eye is more

critical of blurring (of any kind) than other types of degradations which introduce some high frequency features which are not present in the original image. None of the observers selected those images for which the image blur is the dominant degradation term. In particular, the sample-frequency ripple effect (which manifests itself as a fine wire mesh over the images) helped to a certain extent to create an illusion of feature (edge) sharpness that made the observers select images with a moderate amount of this effect as the "best" images.

The subjective evaluation suggests that to the untrained human eye, image blur (i.e. any supression of high frequencies) is often more annoying than sampling artifacts which may create an illusion of sharpness. However, much work still needs to be done. In planned extensions, we will control viewing conditions more stringently to eradicate some biases that may be reflected in our current results. We plan to use a digital monitor instead of film since we have experienced a great deal of contrast and texture variability with film. There is also the need for more exhaustive testing (more scenes with a greater variation in the frequency content etc.) under more controlled conditions. The end-to-end model must be improved to incorporate more sophisticated models of the acquisition and display subsystems as well as physcophysical parameters such as the *contrast sensitivity function.* The sophistication of the human subjects with respect to digital image processing fundamentals may also be a significant bias factor when testing certain images. Finally, the whole experiment would be incomplete unless an end-to-end (initial scene to the reconstructed scene) simulation is performed and the analytical results are correlated with visual testing.

Acknowledgments

The authors would like to thank several people who helped put this experiment together. In particular, they would like to express their gratitude to Steve Reichenbach for his help with the simulation software and Kathy Stacy, Des Leonard and Betsy Avis at NASA Langley Research Center for their help in processing the images. They also thank the graduate students in the Computer Science department of the College of William & Mary who helped with the subjective evaluations of the test images.

References

[1] Park S.K and Schowengerdt R.A, 1982 *Appl. Optics*,**21**,3142

[2] Cook R.L, *ACM Trans. Graphics*, Vol. 5 No. 1 Jan, 1986.

[3] Mitchell, D.P and Netravelli A, *ACM Trans. Graphics* Vol. 22 Aug, 1988.

[4] Schreiber W.F, *Fundamentals of Electronic Imaging Systems* Springer-Verlag, 1986.

[5] Harris F.C, *Proc. IEEE*, **66**, No. 1, Jan. 1978.

[6] Brown E. F, *J. of the SMPTE*, Vol 78, April 1969

[7] Schreiber W.F, Troxel D.E, *IEEE Trans. on PAMI*, PAMI-7, No. 2, March 1985

[8] Schowengerdt R.A, Park S.K and Gray R, 1984 *Int. J. Remote Sensing*, Vol 5. No. 2.

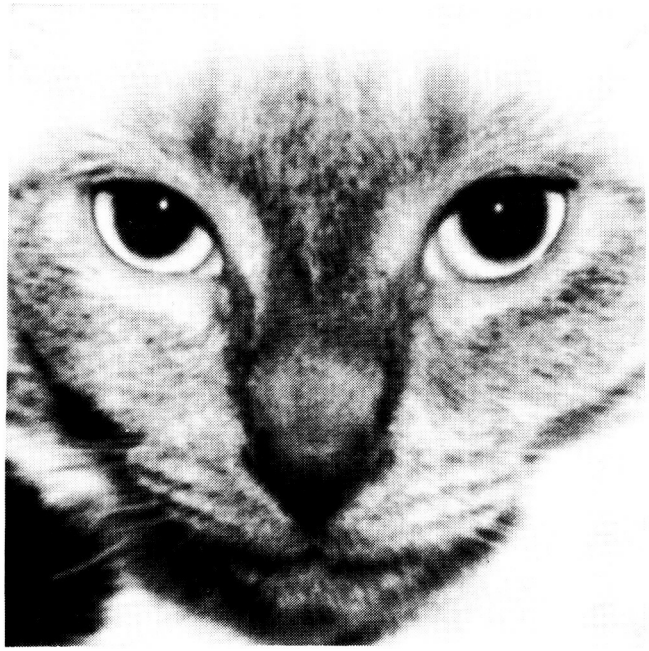

Fig. 5 : Original CAT image with
CTF FACTOR=0.6 and no SR degradation.

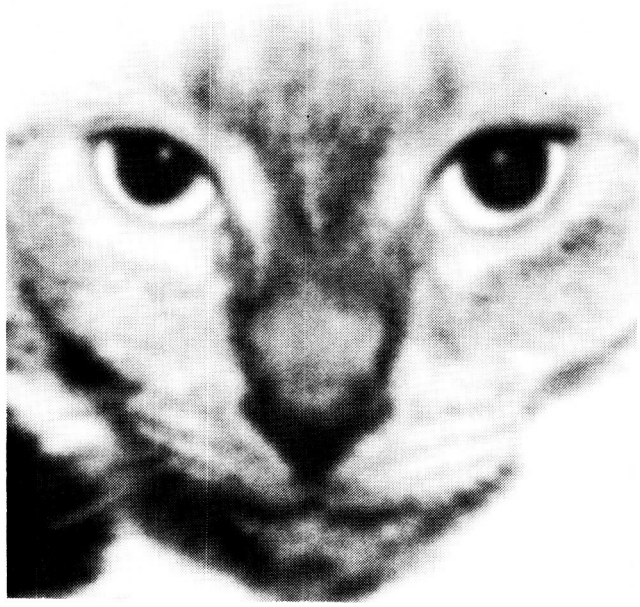

Fig 6 : Reconstructed scene with
CTF FACTOR=0.6 and DTF FACTOR=0.3

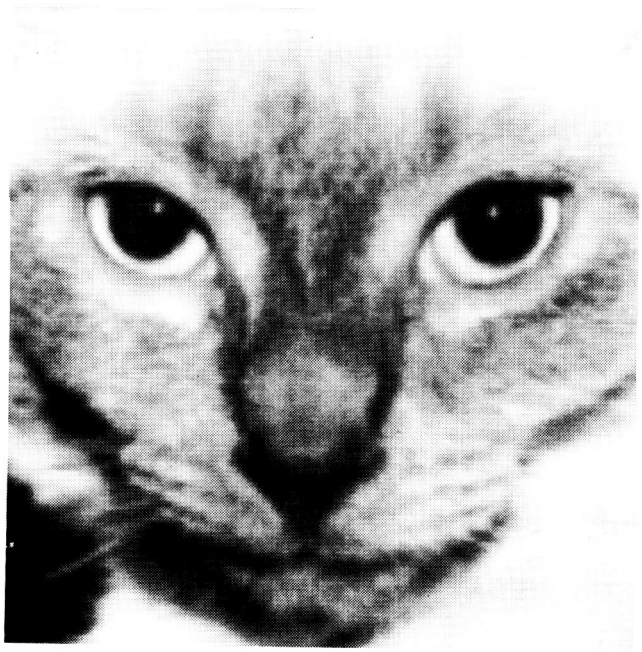

Fig 7 : Reconstructed scene with
CTF FACTOR = 0.6 and DTF FACTOR=0.6
(chosen as the best reconstruction of
Fig 5. by selected observers)

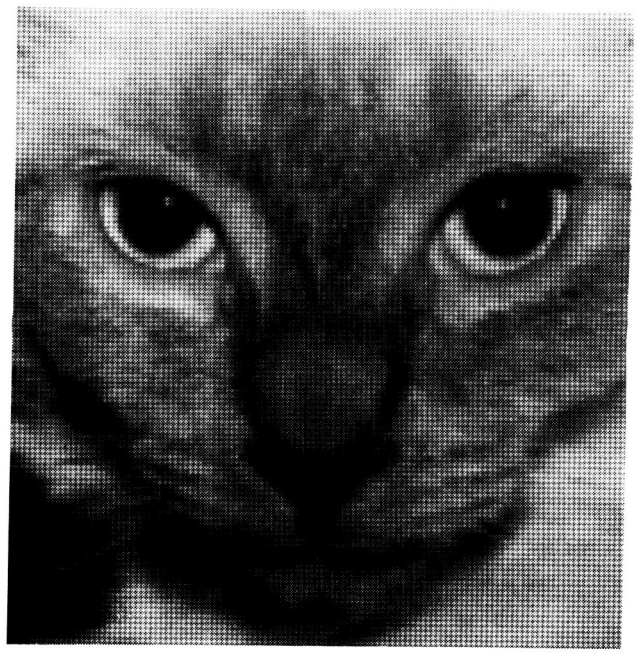

Fig 8 : Reconstructed scene with
CTF FACTOR=0.6 and DTF FACTOR=0.7

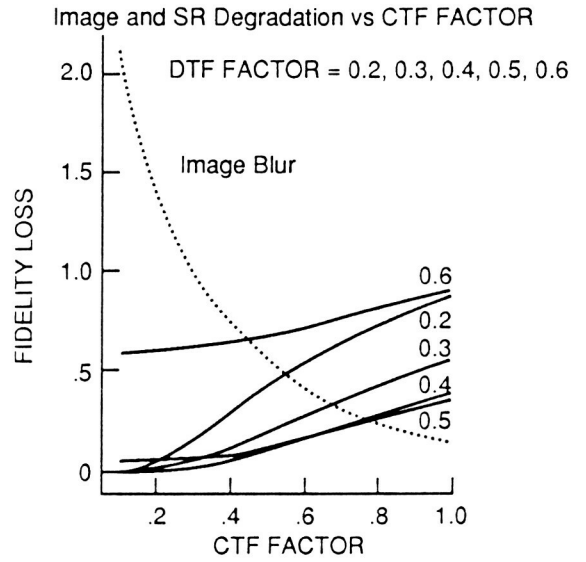

Figure 9: Image and SR Blur vs CTF FACTOR (Dollar)

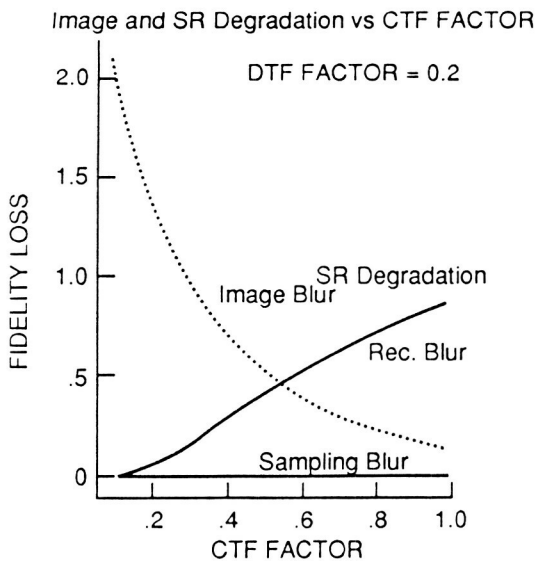

Figure 10: Image and SR Blur vs CTF FACTOR (Dollar)

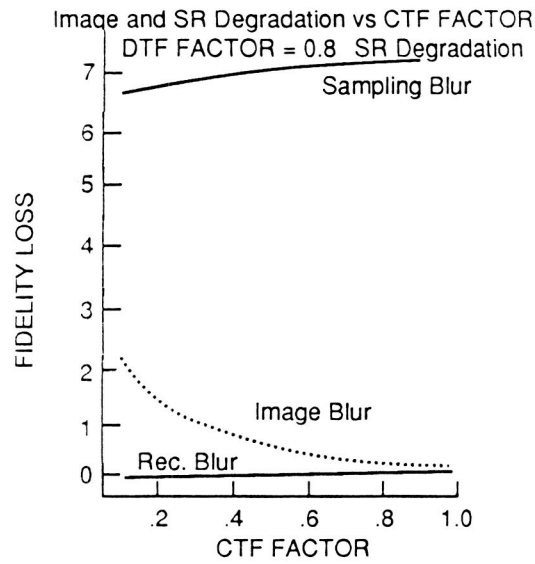

Figure 11: Image and SR Blur vs CTF FACTOR (Dollar)